

Research Article

A Combined Prediction Model for Hog Futures Prices Based on WOA-LightGBM-CEEMDAN

Xiang Wang ¹, Shen Gao,² Yibin Guo,¹ Shiyu Zhou,³ Yonghui Duan,² and Daqing Wu⁴

¹Department of Civil Engineering, Zhengzhou University of Aeronautics, No. 15, Wenyuan West Road, Zhengdong New District, Zhengzhou 450015, China

²Department of Civil Engineering, Henan University of Technology, No. 100, Lianhua Street, Gaoxin District, Zhengzhou 450001, China

³Department of Management, Shanghai University, No. 99 Shangda Road, Baoshan District, Shanghai 200444, China

⁴Department of Economics and Management, Shanghai Ocean University, No. 999, Huchenghuan Road, Pudong New District, Shanghai 201306, China

Correspondence should be addressed to Xiang Wang; shanghaiwx1976@126.com

Received 8 September 2021; Revised 14 January 2022; Accepted 3 February 2022; Published 27 February 2022

Academic Editor: Ning Cai

Copyright © 2022 Xiang Wang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

An integrated hog futures price forecasting model based on whale optimization algorithm (WOA), LightGBM, and Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN) is proposed to overcome the limitations of a single machine learning model with low prediction accuracy and insufficient model stability. The simulation process begins with a grey correlation analysis of the hog futures price index system in order to identify influencing factors; after that, the WOA-LightGBM model is developed, and the WOA algorithm is used to optimize the LightGBM model parameters; and, finally, the residual sequence is decomposed and corrected by using the CEEMDAN method to build a combined WOA-LightGBM-CEEMDAN model. Furthermore, it is used for comparison experiments to check the validity of the model by using data from CSI 300 stock index futures. Based on all experimental results, the proposed combined model shows the highest prediction accuracy, surpassing the comparative model. The model proposed in this study is accurate enough to meet the forecasting accuracy requirements and provides an effective method for forecasting future prices.

1. Introduction

China's futures market has grown rapidly in recent years. A commodity future is a contract in which a certain number of commodities will be delivered by an exchange at a future time. As a form of risky investment and risky return, futures trading is also a very important investment tool for investors. Meanwhile, the futures market can reasonably use and gather a vast amount of social idle capital, which is valuable to China's market economy. For both companies and investors, an accurate prediction of prices in the futures market is a key guide. Due to the fact that commodity futures are influenced by a variety of factors, which can cause large fluctuations in price, it is difficult to achieve accurate price

control. Therefore, the accurate prediction of futures prices has become a hot research topic.

In the research on price forecasting in futures markets, traditional econometric models are used along with machine learning models. Futures prices are used as time series in most econometric models, which employ statistical methods to make linear forecasts. Common methods include autoregressive integrated moving average (ARIMA) [1, 2] and generalized autoregressive conditional heteroscedasticity (GARCH) [3, 4]. In contrast, the prices in the futures market are often influenced by a variety of factors along with the characteristics of nonstationarity, nonlinearity, and high complexity, which can lead to large errors when only making linear forecast [5, 6].

By contrast with econometric models, machine learning models are effective in mining and retaining the valuable information in the data and in dealing with nonlinear data effectively [7]. Artificial neural network (ANN) [8, 9], support vector machine (SVM) [10, 11], long short-term memory (LSTM) [12], and ensemble learning models are some of the common models. Ensemble learning models can effectively combine the results of multiple base learners to achieve secondary learning of the problem with high generalizability. The ensemble learning model uses boosting to train the base learner by serial learning. This strategy reduces the prediction bias of the model and improves the algorithm's ability to fit. Zhang and Hamori [13] used extreme gradient boosting (XGBoost) as an experimental model for crude oil futures price forecasting, and the results showed that the XGBoost model was able to achieve an accuracy rate of 86%. Deng et al. [14] used XGBoost model to predict the price of apple futures, and bagging ensemble learning was used to further integrate and optimize the model in order to reduce overfitting. To predict the LME nickel settlement price, Gu et al. [5] employed the empirical wavelet transform (EWT) and gradient boosting decision tree (GBDT). Luo et al. [15] used genetic algorithm (GA) to optimize the parameters of their GBDT model and predict copper and soybean futures prices in China, which proved to be superior to BP neural network and SVM models. In spite of the above boosting algorithm's advantages, there are some problems; for example, the GBDT model takes too long to diagnose when processing the data of complex samples, resulting in low prediction efficiency [16]; the XGBoost model still traverses the data set during the node splitting process, increasing the computational burden. LightGBM is an improved version of the GBDT model using a unique leaf-wise growth strategy based on the maximum depth limit, which can reduce more errors and get better accuracy while using the same number of splits. While the histogram algorithm of the LightGBM can improve model running efficiency while reducing memory footprint, it has not been applied to the futures price forecasting problem as it has been used in several fields [17–19].

In order to further improve machine learning models' accuracy, most studies focus on the models themselves or data but ignore the valuable information hidden in the sequence of prediction residuals. The information from these hidden residual sequences can significantly enhance the final prediction [20]. Usually, residual series resulting from forecasting are a kind of time series with nonpure randomness and autocorrelation, and the decomposition-individual forecasting-ensemble method is a desirable way to handle such characteristics. The existing studies include many decomposition methods, including variational mode decomposition (VMD) [21, 22], wavelet transform (WT) [23], and empirical mode decomposition (EMD). Among them, the EMD method can decompose data into multiple intrinsic mode functions (IMF) according to the data characteristics, which has a better ability to decompose for nonlinear and nonsmooth data and can effectively extract the characteristics of the data at different frequency scales [24]. EMD, however, is prone to the phenomenon of modal

mixing during decomposition, which affects its decomposition performance. Wu [25] proposed ensemble empirical mode decomposition (EEMD) to improve the EMD, which can effectively solve the modal mixing problem by adding Gaussian white noise to the original signal, but there is still residual Gaussian white noise in the decomposed IMFs, resulting in incorrect reconstruction. To improve EEMD, Torres proposed the CEEMDAN method [26]. CEEMDAN adds adaptive white noise at each stage, which effectively overcomes EEMD's large reconstruction error. CEEMDAN has been used in several areas of forecasting because of its advantages. Zhang et al. [27] applied the CEEMDAN method to decompose wind speed series and used a neural network model to predict each IMF component, and finally the prediction results of each component were combined, and the experimental results showed that the method could effectively improve forecast accuracy. Wang et al. [28] combined CEEMDAN decomposition method and GRU neural network to predict natural gas price. The CEEMDAN decomposition method was used by Zhao and Chen [29] to decompose the carbon price, and the extreme learning machine (ELM) model was optimized with the improved sparrow algorithm to forecast each IMF component, and the results showed that the combined approach can effectively improve the forecasting accuracy. In their study, Cao et al. [30] established an EMD-LSTM model and a CEEMDAN-LSTM model to forecast stock market prices, and, based on empirical analysis, the CEEMDAN-LSTM model had more accurate predictions.

Following the above analysis, this paper uses the LightGBM model to forecast the futures price and the whale optimization algorithm (WOA) to select the hyperparameters of the model. For further improvement of the model prediction results, the CEEMDAN decomposition method is used to decompose the residual series of the predictions of LightGBM. As the support vector regression (SVR) model has better nonlinear fitting and generalization ability, it is used to predict each component generated by CEEMDAN, and the results are combined after the prediction is completed, resulting in a combined WOA-LightGBM-CEEMDAN model. A combined forecasting model is used to forecast the price of hog futures in China. We selected hog futures prices as the subject of this paper for two main reasons. Firstly, hogs are one of the most important agricultural products in China, and they provide a significant percentage of the country's meat consumption. According to data published by the National Bureau of Statistics, pork accounted for more than 75% of China's total meat consumption from 2013 to 2019 [31]. In any case, the dramatic fluctuations in the price of pork have had a significant impact on both the balance of supply and demand on the market as well as on both farmers and consumers. In the futures market, there are functions of price discovery and hedge, which can mitigate the economic loss caused by price fluctuation, bring income to investors, and help farmers to adjust the scale of pig breeding appropriately so that economic benefits are maximized. As a result, it is imperative to produce accurate forecasts for hog futures prices in order to stabilize

hog market prices and maintain a balance between supply and demand. Furthermore, in comparison to other types of price data, the price of hog futures is influenced to a greater degree by market supply and demand and is relatively less influenced by government macrocontrol, while the price data is to some degree influenced by the cycle. Finally, on January 8, 2021, hog futures will be listed and traded in mainland China. Currently, there is less discussion on predicting hog futures, and the model in this study is used to predict hog futures prices in China, which has some significance for future research of the same type.

The following are the main steps in this paper for forecasting the hog futures price. To begin with, we establish a system of hog futures price indexes and employ a grey correlation analysis to identify the main factors affecting the futures price of hogs, thereby improving the model's prediction accuracy. Additionally, LightGBM is used to establish a hog futures price forecasting model, and WOA is used to optimize model parameters in order to eliminate forecasting errors caused by the parameter settings of LightGBM. Furthermore, in order to improve the model prediction accuracy, the CEEMDAN method is used to correct the residual series of LightGBM prediction results in order to construct a combined WOA-LightGBM-CEEMDAN model.

The remainder of this paper is organized as follows. The second section provides an introduction to the LightGBM model, WOA, and CEEMDAN, followed by a description of the implementation steps of the combination model in this paper. In Section 3, we describe the prediction index system and data used in this paper and provide the parameter settings for the model. Our experimental analysis and discussion of hog futures prices are presented in Section 4. In Section 5, we summarize some conclusions and suggest directions for future research.

2. Materials and Methods

2.1. LightGBM. The LightGBM model, developed by Microsoft, is an open-source gradient boosting model based on decision trees. The LightGBM model is also capable of parallel learning, similar to the XGBoost model. LightGBM, however, has the advantage of a faster training rate and less memory consumption compared to XGBoost [32].

Consider a set of data sets $M = \{x_i, y_i\}_1^N$, in which $x = \{x_1, \dots, x_n\}$ is the input to the model and y is the prediction label. The model function is $F(x)$ and the loss function is $L(y, F(x))$. In gradient boosting, the negative gradient of the loss function L is used instead of the residuals to determine the value of the current model function $F(x)$. Taking g_{ij} as the negative gradient of the j th iteration, we obtain

$$g_{ij} = - \left[\frac{\partial L(y_i, F(x_i))}{\partial F(x_i)} \right]_{F(x)=F_{j-1}(x)}. \quad (1)$$

If $h(x)$ is the weak learner, then $h(x)$ should be used to fit the negative gradient of the loss function to find the best fit value as follows:

$$g_j = \arg \min_g L(y_i, F_{j-1}(x_i) + gh_j(x_i)). \quad (2)$$

The model update formula is defined as follows:

$$F_j(x) = F_{j-1}(x) + g_j h_j(x). \quad (3)$$

In the above approach, gradient boosting is updated iteratively; one weak learner is trained at a time. After the iterations are completed, the weak learners are added together to obtain the strong learner.

To accelerate the training of the gradient boosting framework model without compromising accuracy, the LightGBM model uses a number of optimization methods, the most prominent of which is the histogram algorithm as well as the leaf-wise growth strategy with depth constraints.

In the LightGBM model, the histogram algorithm is a method of discretizing data, which reduces the computational cost and memory consumption, thus improving its efficiency.

The decision tree growth strategy used in the traditional gradient boosting framework model is a very inefficient layer-by-layer growth strategy, since it treats the subleaves of the same layer in an indiscriminate way, causing unnecessary model runs and thus increasing the burden on the model. Leaf-wise growth strategy with depth limit finds the leaf with the highest splitting gain from all the current leaves, then splits, and so forth. With the same number of splits as the layer-by-layer growth strategy, the leaf-wise growth strategy can effectively reduce errors and increase prediction accuracy. In addition, LightGBM includes a maximum depth limit that enables the model to achieve maximum prediction accuracy while preventing overfitting.

2.2. Whale Optimization Algorithm. The whale optimization algorithm [33] (WOA) is a population intelligence algorithm developed by Australian scientist Mirjalili and Lewis in 2016. The purpose of the algorithm is to determine optimal target parameters by simulating the feeding behavior of humpback whales. In the WOA algorithm, the location of each humpback whale represents a viable solution for a set of parameters, and changes in location are made in three different ways: encircling prey, spiral search, and random search.

2.2.1. Encircling Prey. The whales first share the information about the location of the searched prey as a group, with the location of the prey or the closest whale to the prey regarded as the optimal solution, and then they approach the whale that is currently closest to the prey's location, thus contracting its encirclement. This behavior is defined as follows:

$$\begin{cases} X(t+1) = X^*(t) - A \cdot D, \\ D = |C \cdot X^*(t) - X(t)|. \end{cases} \quad (4)$$

The position of the killer whale is X ; $X^*(t)$ is the current optimal position of the whale; t is the current number of iterations; A and C are the coefficient matrices, and the expressions are

$$\begin{cases} A = 2a \cdot r_1 - a, \\ C = 2r_2, \\ a = 2 - \frac{2t}{t_{\max}}. \end{cases} \quad (5)$$

There are two random numbers, r_1 and r_2 , taking values ranging from 0 to 1; a is the convergence factor, linearly decreasing from 2 to 0; and t_{\max} is the global maximum number of iterations.

2.2.2. Spiral Search. In the search phase, the whale approaches its prey along an ascending spiral, and the mathematical expression for this phase is

$$\begin{cases} X(t+1) = X^*(t) + D \cdot e^{bl} \cos(2\pi l), \\ D = |C \cdot X^*(t) - X(t)|. \end{cases} \quad (6)$$

In this scenario, b is a constant parameter determining the shape of the spiral, and l is a uniformly distributed random number varying from -1 to 1 .

During rotational search, the whale contracts its envelope to approach its prey. By assuming that each has a 50% probability of rotational search and envelope contraction, the expression can be defined as follows:

$$X(t+1) = \begin{cases} X^*(t) - A \cdot D, & p < 0.5, \\ X^*(t) + D \cdot e^{bl} \cos(2\pi l), & p \geq 0.5, \end{cases} \quad (7)$$

where p is a random number with values ranging from 0 to 1.

2.2.3. Random Search. For the purpose of enhancing the global search capability of the whale, WOA has also developed a random search algorithm to further increase the search range. When $|A| \geq 1$, the whale is outside the envelope and the whale moves away from the current optimal solution and performs a random search. Conversely, when $|A| < 1$, a spiral search is used to update the position. The expressions can be defined as follows:

$$\begin{cases} X(t+1) = X_{\text{rand}}(t) - A \cdot D^*, \\ D^* = |CX_{\text{rand}}(t) - X(t)|, \end{cases} \quad (8)$$

where $X_{\text{rand}}(t)$ is the random position of the whale.

2.3. WOA-LightGBM. Model parameters have a considerable influence on model prediction effects, and there are many parameters in the LightGBM model which affect the model in different ways. Therefore, this paper borrows from the previous study [34] to set the LightGBM parameters to be searched for, that is, the number of boosted trees to fit ($n_{\text{estimators}}$), the learning rate (learning_rate), maximum tree depth for base learners (max_depth), and maximum tree leaves for base learners (num_leaves). By using the parameters of the LightGBM model as the position vectors of each whale, the WOA seeks the global optimal position in

the algorithm through iterative search and outputs it as the final parameters of the LightGBM model. The specific processes are as follows:

Step 1: Initialize the whale optimization algorithm. Set the number of whale populations, the maximum number of iterations, and the whale search area boundaries for the WOA algorithm.

Step 2: Set up the fitness function. In the WOA algorithm, first, the position of the current population is randomly initialized within the boundary range, then the fitness function is used to calculate the fitness value for each whale in the current population, and then the whale with the smallest fitness value is chosen as the global optimal solution for the current population. In the LightGBM model, the fitness function is selected as the mean square error function, and the specific expression is as follows:

$$f(x_i) = \frac{1}{n} \sum_{j=1}^n (\theta_{i,j} - \hat{\theta}_{i,j})^2, \quad (9)$$

where x_i is the location of the i th individual whale, $\theta_{i,j}$ is the corresponding j th true value in the i th individual, and $\hat{\theta}_{i,j}$ is the predicted value derived from the LightGBM model based on the parameters set for x_i .

Step 3: Maintain the position of individual whales according to the three methods of encircling prey, spiral search, and random search within the WOA, and control all whales within a predetermined boundary.

Step 4: Upon completion of the position update, the whale position is input into the fitness function in order to calculate the fitness result, and the optimal whale position is selected as the current global optimal solution.

Step 5: Repeat Step 3 and Step 4 until the algorithm has reached the maximum number of iterations.

Step 6: Output the final WOA algorithm search results and integrate them into the LightGBM model for modeling predictions.

The flow chart of the above steps is depicted in Figure 1.

2.4. CEEMDAN Residual Correction Model. Despite the fact that the LightGBM model for hog futures price forecasting can utilize historical time data to obtain better results, there are still a number of residual series that display nonlinearity and large degrees of randomness. Therefore, a method for forecasting and correcting the residual series is needed to improve the model's forecasting accuracy.

EMD is a technique for decomposing nonlinear, non-smooth sequences into IMFs components with different fluctuation scales; however, it is susceptible to modal mixing when decomposing the signals, which interferes with the decomposition process. EEMD can effectively resolve the modal mixing phenomenon by adding Gaussian white noise to the original signal; however, there is still Gaussian white noise present in the components of the IMF decomposed by

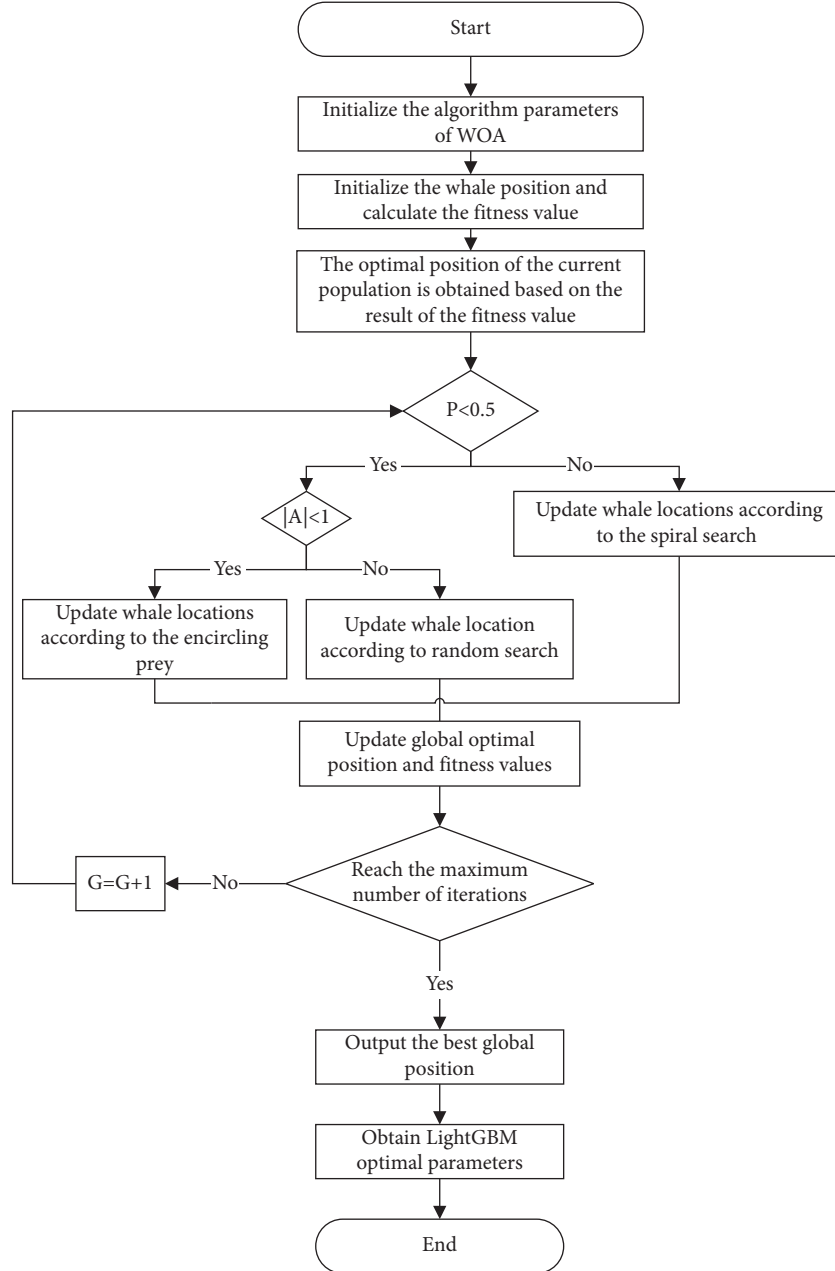


FIGURE 1: Flow chart of WOA-LightGBM model pred.

the EEMD method, which can lead to errors during reconstruction. Due to this, the CEEMDAN [26] method proposed by Torres incorporates adaptive white noise at each stage in order to overcome the large reconstruction error of the EEMD method. Therefore, the CEEMDAN method is employed in this study in order to decompose the residual series and predict each IMFs component as well as the trend term component separately.

CEEMDAN can be decomposed into the following steps:

Step 1: Obtain the residual series. The WOA-LightGBM model is used to model and forecast the hog futures price, and the residual series is obtained by calculating

the difference between the forecast results and the true values:

$$y(t) = Y_{\text{real}} - Y_{pr}, \quad (10)$$

where $y(t)$ symbolizes the residual series; Y_{real} is the true value of hog futures; and Y_{pr} is the forecast value derived by the WOA-LightGBM model.

Step 2: An adaptive Gaussian white noise sequence is added to the original sequence to obtain the new sequence with noise:

$$\bar{y}_i(t) = y(t) + \sigma n_i(t), \quad i = 1, 2, \dots, N. \quad (11)$$

$y(t)$ represents the original residual sequence; $\bar{y}_i(t)$ represents the new sequence with white noise added to it; $n_i(t)$ denotes the white noise added to the original data; σ denotes the adaptive coefficient.

Step 3: On the new sequence M , the EMD decomposition is applied after the addition of white noise to obtain N IMFs, and the first IMF of CEEMDAN can be derived by averaging the N modal components as follows:

$$\text{imf}_1(t) = \frac{1}{N} \sum_{i=1}^N \text{imf}_{1i}(t). \quad (12)$$

Hence, $R_1(t)$ is the residual component at this point:

$$R_1(t) = \bar{y}_i(t) - \text{imf}_1'(t). \quad (13)$$

Step 4: The adaptive white noise sequence $\sigma n_i(t)$ is added to $R_1(t)$ to form a new sequence with noise $R_1(t) + \sigma E_1(n_i(t))$, where $E_j(\bullet)$ is the j th eigenmodal component obtained from the EMD decomposition. By decomposing the new series with the EMD and averaging, the second IMF and the residual component are obtained.

$$\text{imf}_2(t) = \frac{1}{N} \sum_{i=1}^N E_1(R_1(t) + \sigma_1 E_1(n_i(t))), \quad (14)$$

$$R_2(t) = R_1(t) - \text{imf}_2'(t).$$

Step 5: Repeat the three preceding steps in order to obtain the $j+1$ th IMF and the j th residual component.

$$\text{imf}_{j+1}(t) = \frac{1}{N} \sum_{i=1}^N E_1(R_j(t) + \sigma_j E_j(n_i(t))), \quad (15)$$

$$R_j(t) = R_{j-1}(t) - \text{imf}_j'(t).$$

Step 6: This procedure is repeated until the CEEMDAN is terminated when the remaining components cannot be decomposed by EMD, and, finally, the original sequence $y(t)$ is decomposed into multiple IMFs and a residual component.

$$y(t) = \sum \text{imf}(t) + R_{es}(t). \quad (16)$$

Once CEEMDAN has decomposed the original series, all IMFs and residual components are predicted separately using appropriate prediction models, and then the individual predictions are linearly combined to determine the final residual prediction. With regard to prediction methods, SVR is an algorithm used for regression modeling by SVM, which has a strong ability to generalize to data that are nonlinear and have stochastic fluctuations [35]. Therefore, in this study, SVR is used to predict each IMF as well as the residual components generated by CEEMDAN separately.

2.5. *WOA-LightGBM-CEEMDAN*. The aim of this study is to improve the forecasting performance of the hog futures price by combining WOA, LightGBM, and CEEMDAN models to formulate the WOA-LightGBM-CEEMDAN model. Here are the specific steps for implementation:

Step 1: Data preprocessing. Preprocess the data and divide it into training sets and testing sets.

Step 2: Obtain the preliminary fitted values. Establish the LightGBM model for forecasting the hog futures price, employ the WOA algorithm to determine the parameters of the model, and then use the optimized model to estimate the preliminary price of the hog futures and obtain the preliminary fitted value.

Step 3: Obtain the residual series. The preliminary fitted value results are compared with the original data, and residual series are derived. The specific formula is as follows:

$$\Delta(t_i) = y_{\text{real}}(t_i) - \hat{y}_{\text{LGB}}(t_i), \quad i = 1, 2, \dots, N, \quad (17)$$

where $\Delta(t_i)$ represents the residual series value; $y_{\text{real}}(t_i)$ represents the true value; and $\hat{y}_{\text{LGB}}(t_i)$ represents the prediction result of the WOA-LightGBM model.

Step 4: The residual series are brought into CEEMDAN for modal decomposition in order to obtain the IMF components of different frequencies with number n , as well as a trend component.

Step 5: The SVR model is used to predict each component of CEEMDAN separately, and the results are combined to obtain the final residual prediction value $\Delta(t_i)$ after the prediction has been completed.

Step 6: Residual correction is performed on the WOA-LightGBM model. After combining the residual prediction results from the WOA-LightGBM model and the preliminary fitting results obtained from the WOA-LightGBM model, we obtain residual correction results for the combined final model:

$$\hat{y}(t_i) = \Delta(t_i) + \hat{y}_{\text{LGB}}(t_i), \quad i = 1, 2, \dots, N. \quad (18)$$

An intuitive implementation flow chart is shown in Figure 2.

3. Data Description

3.1. *Impact Factors of Hog Futures*. Hog futures prices are influenced by a variety of factors. The purpose of this paper is to identify primary indicators of hog futures price from three perspectives: supply, demand, and futures market.

Among the supply factors, the price of piglets is an important input before pig slaughtering, and changes in piglet prices will directly affect the cost of production of pigs. The price of sows directly influences the number and price of piglets, which impacts the cost of pig breeding. Feed is another important input for hog production, and changes in feed prices will have an impact on the size of production.

At the level of demand factors, when the price of pork exceeds consumers' psychological expectations, they will

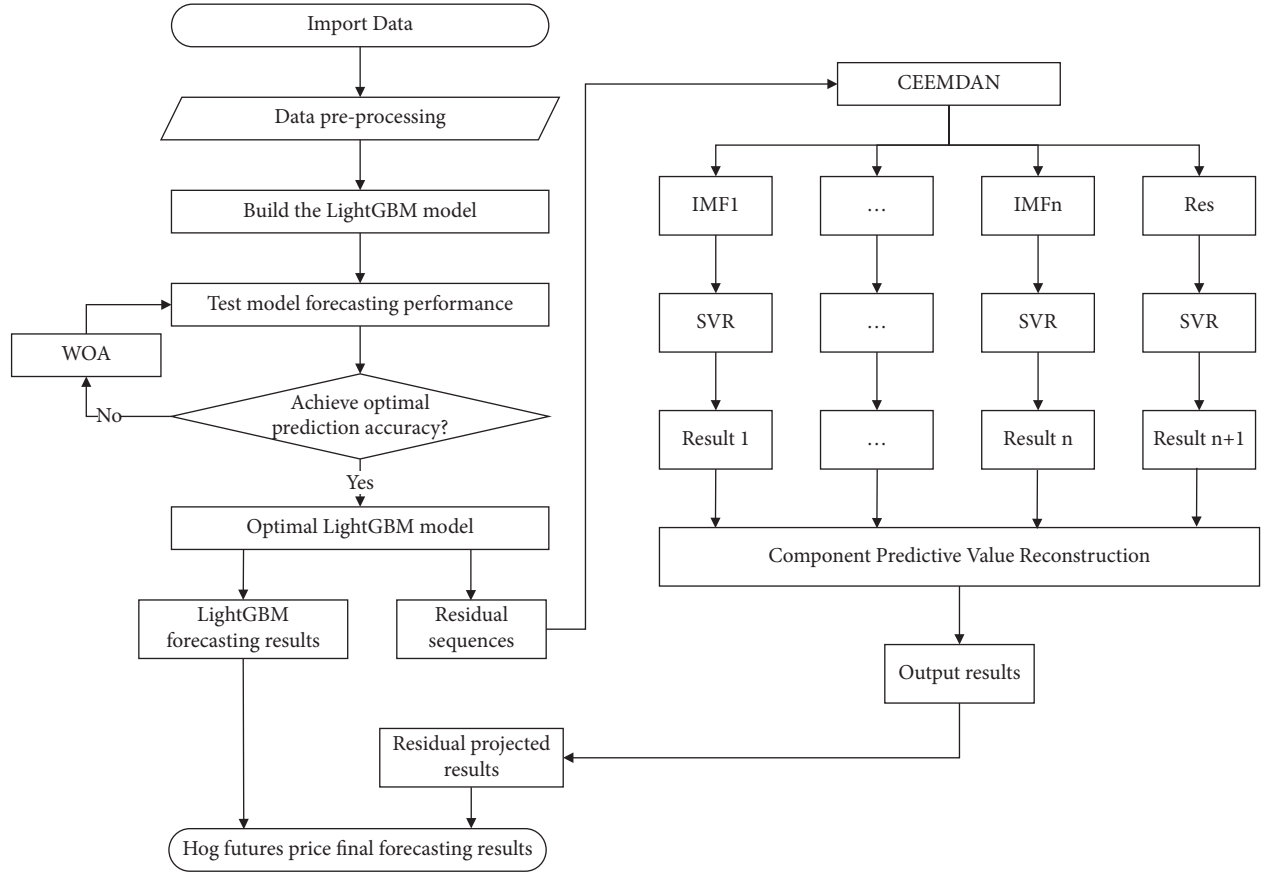


FIGURE 2: Structure flow chart of WOA-LightGBM-CEEMDAN.

prefer alternatives with lower prices, so the price of alternatives will directly affect the consumption of pork and therefore the price of hogs.

At the level of futures market factors, volume and open interest can accurately reflect the relationship between supply and demand as well as the volatility of the futures market, while providing helpful information in predicting the overall trend of the market. Basis is the difference between the spot price and the futures price and is a dynamic indicator of the actual change between the two prices; basis changes directly impact the effectiveness of the hedge. The spot price is the foundation of the futures price. The spot price always appears prior to the futures price, while, at the same time, the delivery price of the futures price is always based on the spot transaction price.

The sow price considered in this study is the national binary 50 kg sow price; the corn and soybean meal prices are used to represent feed prices; and the beef price and lamb price are used to represent alternative prices. Table 1 provides detailed indicators for each classification.

3.2. Grey Relation Analysis. Grey relation analysis (GRA) is used to further screen the indicators in order to improve their accuracy and reduce their error. The GRA [36] method measures the degree of association between the reference and comparison series through grey correlation. The specific calculation process is as follows:

Step 1: Take the hog futures price data $Y_i = \{y_{i1}, y_{i2}, \dots, y_{in}\}$ ($i = 1, 2, \dots, n$) as the reference series and the influencing factors $X_j = \{x_{j1}, x_{j2}, \dots, x_{jn}\}$ ($j = 1, 2, \dots, n$) in Table 1 as the comparison series.

Step 2: Calculate the grey correlation coefficients for each reference sequence and comparison sequence. The calculation formula is

$$\xi_{ij}(k) = \frac{\min_i \min_k |x_j(k) - y_i(k)| + \rho \cdot \max_i \max_k |x_j(k) - y_i(k)|}{|x_j(k) - y_i(k)| + \rho \cdot \max_i \max_k |x_j(k) - y_i(k)|}. \quad (19)$$

ρ is a resolution factor that takes values between 0 and 1. For this paper, we take $\rho = 0.5$; $i = 1, 2, \dots, n$, $j = 1, 2, \dots, n$; $k = 1, 2, \dots, n$.

Step 3: Calculate the correlation:

$$r_{ij} = \frac{1}{n} \sum_{k=1}^n \xi_{ij}(k). \quad (20)$$

r_{ij} is the correlation index value between the reference series and the comparison series. An index value close to 1 indicates a higher degree of correlation between the comparison series and the reference series.

Table 2 shows the GRA results for hog futures price influencing factors. Among the many indicators, the factors with grey correlations less than 0.7 were removed in this paper, and the ultimate hog futures price influencing factors indicator system is shown in Table 3.

3.3. Data Source. This paper forecasts hog futures prices and establishes the influencing factors from three perspectives: supply, demand, and futures markets. Since hog futures are listed and traded on the Dalian Commodity Exchange on January 8, 2021, the data time points are all selected as daily data from January 8, 2021, to July 22, 2021, for a total of 129 samples for simulation experiments. In the experiments, 70% of the data are used as training data and 30% as testing data. Among them, except for lamb price and beef price from Wind database, other indicators are published on Huarong Rongda data analyst website (<https://dt.hrrdqh.com/>). Based on the above data, piglet price and sow price are weekly data, which were converted into daily data by using the EVIEWS software. Table 4 presents descriptive statistics for each indicator.

3.4. Experiment Preparation

3.4.1. Model Evaluation. The mean square error (MSE), the mean absolute error (MAE), the root mean square percentage error (RMSPE), the mean absolute percentage error (MAPE), R -square (R^2), and directional accuracy (DA) [37] are selected as the model evaluation functions. In general, when the values of MSE, MAE, RMSPE, and MAPE indicators are smaller and the values of R^2 and DA indicators are larger, the model predicts better and vice versa. Here is the formula for the functions:

$$\begin{aligned}
 \text{MSE} &= \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2, \\
 \text{MAE} &= \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|, \\
 \text{RMSPE} &= \sqrt{\frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right|^2} * 100\%, \\
 \text{MAPE} &= \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| * 100\%, \\
 R^2 &= 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y}_i)^2}, \\
 \text{DA} &= \frac{1}{N} \sum_{i=1}^N a(i) \times 100\%.
 \end{aligned} \tag{21}$$

In the above formula, N denotes the number of samples in the test set; y_i is the true value; \hat{y}_i is the forecasting result; when $(y(i+1) - y(i))(\hat{y}(i+1) - y(i)) \geq 0$, $a(i) = 1$; otherwise, $a(i) = 0$.

3.4.2. Data Preprocessing. Based on the descriptive statistics shown in Table 4, it appears that the order of magnitude of the factors varies due to the different units. To avoid prediction errors due to the order of magnitude of the data, all predicted data are normalized by using the following equation:

$$x^* = \frac{x - x_{\min}}{x_{\max} - x_{\min}}, \tag{22}$$

where x^* is the normalized data value; x is the input data, and x_{\min} and x_{\max} are the minimum and maximum values of the input data.

3.4.3. Model Parameters Setting. To analyze and compare the model prediction performance, multiple algorithms were used to measure the prediction effect of the WOA-LightGBM-CEEMDAN model. For the single model, SVR, BPNN, extreme learning machine (ELM), GBDT, and XGBoost models are selected to compare and analyze the prediction performance of LightGBM. In terms of WOA algorithm performance, the grey wolf optimization (GWO) algorithm is selected to optimize the LightGBM model (GWO-LightGBM) to analyze and compare the WOA-LightGBM. As to the prediction effect of the residual correction combination model, WOA-LightGBM-EEMD, which uses EEMD method to decompose the residual series, and WOA-LightGBM-SVR, which directly performs SVR model prediction without decomposing the residual series first, are used as comparative analysis models.

For the above model, given that GBDT and XGBoost have many parameters, the model parameters selection of literature [34] is used to select parameter values. We select the number of boosting stages to perform ($n_{\text{estimators}}$), boosting learning rate (learning_rate), the minimum number of samples required to split an internal node (min_samples_split), and the minimum number of samples required to be at a leaf node (min_samples_leaf) in the GBDT model as the parameters to be searched for. In the XGBoost model, number of gradient boosted trees ($n_{\text{estimators}}$), boosting learning rate (learning_rate), maximum tree depth for base learners (max_depth), subsample ratio of the training instance (subsample), and subsample ratio of columns when constructing each tree (colsample_bytree) are selected as the parameters to be searched for.

The specific parameters of each model are shown in Table 5.

4. Results Discussion

4.1. Analysis of the Preliminary Fitting Performance of WOA-LightGBM. Figure 3 illustrates WOA-LightGBM's preliminary performance on the testing set for predicting hog futures prices. Figure 3 illustrates the fact that while the WOA-LightGBM model can predict the trend of hog futures prices reasonably well, it cannot predict the exact futures prices, and there are still a series of residuals. Therefore, CEEMDAN is used to decompose the residual series generated by WOA-LightGBM.

TABLE 1: Primary indicator system for hog futures prices.

Category	Indicators
Supply	Piglet price
	Sow price
	Corn price
	Soybean meal price
Demand	Lamb price
	Beef price
Futures market	Open interest
	Volume
	Basis
	Spot price
	Previous closing price

TABLE 2: Results of GRA.

Category	Indicators	GRA
Supply	Piglet price	0.85
	Sow price	0.81
	Corn price	0.76
	Soybean meal price	0.64
Demand	Lamb price	0.81
	Beef price	0.77
Futures market	Open interest	0.64
	Volume	0.73
	Basis	0.69
	Spot price	0.79
	Previous closing price	0.96

TABLE 3: Final indicators of hog futures price impact factors.

Category	Indicators
Supply	Piglet price
	Sow price
	Corn price
Demand	Lamb price
	Beef price
Futures market	Volume
	Spot price
	Previous closing price

The specific methods of CEEMDAN are as follows:

Step 1: The WOA-LightGBM model is predicted separately for the training and testing sets to obtain the prediction results for the entire sample length.

Step 2: The WOA-LightGBM model predictions are subtracted from the true value series to obtain the model residual series, and CEEMDAN is used to decompose the residual series into IMFs and residual components.

Step 3: For each component, the SVR model is trained using 70% of the residual series, and the remaining 30% is used for prediction.

Step 4: The prediction results of the SVR model for each component are summed to obtain the final residual prediction results for the test set.

Figure 4 displays the results of the CEEMDAN decomposition of the total sample length, with a total of four IMFs and one residual component.

4.2. *Analysis of Model Prediction Performance.* Figure 5 and Table 6 illustrate the fitting curves of each model for the hog futures prices in the testing set sample and the prediction performance for the six evaluation indicators, respectively. This analysis leads to the following conclusions.

The WOA-LightGBM model is optimal for one-model prediction, which can be attributed primarily to the following factors. Firstly, SVR, ELM, and BPNN models relate to a single machine learning algorithm, while LightGBM belongs to a boosting ensemble learning framework, which can effectively improve the prediction accuracy and generalization ability of the model by combining the predictions of multiple base learner algorithms. Secondly, compared with GBDT and XGBoost, which are part of the decision tree framework, LightGBM's unique leaf-wise subleaf growth strategy may effectively improve the prediction efficiency of the algorithm, while the depth limit may effectively prevent the overfitting problem of the model.

As for the optimization performance of the WOA, the prediction results of the GWO-LightGBM model and the WOA-LightGBM model can be analyzed, and it can be seen that the prediction accuracy of the WOA-LightGBM model is improved by 37.24%, 19.42%, 20.54%, 19.31%, 0.28%, and 5.71% for MSE, MAE, RMSPE, MAPE, R^2 , and DA, respectively. Under the same parameter setting, as compared to GWO algorithm, WOA can lock the optimal parameters of the LightGBM model more quickly and improve the accuracy of the prediction.

In terms of the combined algorithm, all combined models have better prediction results than single models, and the WOA-LightGBM-CEEMDAN model has the highest prediction accuracy. The results demonstrate that the CEEMDAN residual correction combination model proposed in this paper can further improve the prediction accuracy of the WOA-LightGBM model, thereby improving the accuracy of forecasting the price of hog futures.

4.3. *Analysis of Model Prediction Errors for WOA-LightGBM-CEEMDAN.* Further analysis of the prediction performance of the WOA-LightGBM-CEEMDAN model is conducted by examining prediction errors between the prediction results and the real values for the WOA-LightGBM-CEEMDAN, WOA-LightGBM-EEMD, WOA-LightGBM-SVR, and WOA-LightGBM models. The prediction errors of the four models are presented in Table 7 and Figure 6. In the forecasting performance analysis of the testing set in Table 7, WOA-LightGBM-CEEMDAN has the best prediction results, with an average error of -1.35 yuan/ton, while the WOA-LightGBM model without residual correction has the largest error in prediction, reaching an average of 71.65 yuan/ton. This can be attributed primarily to the following factors: (a) Because residual sequences of a single machine learning model may contain valuable information that can boost the final prediction effect, the WOA-LightGBM model

TABLE 4: Descriptive statistics of hog futures price data.

	Count	Mean	Std	Min	25%	50%	75%	Max
Futures price	128.0	24439.2	3760.8	16735.0	20167.5	25412.5	27297.5	29380.0
Piglet price	128.0	251.1	101.5	91.8	154.9	264.0	340.9	382.6
Sow price	128.0	769.3	132.8	469.9	781.8	825.6	851.8	887.2
Corn price	128.0	2822.8	57.3	2527.8	2797.2	2835.6	2862.8	2908.7
Lamb price	128.0	75992.9	1632.1	71930.0	75450.0	76180.0	77280.0	78420.0
Beef price	128.0	77580.5	668.7	76200.0	77087.5	77635.0	78155.0	78930.0
Volume	128.0	12669.2	13361.3	1601.0	3684.0	7750.5	16896.8	91056.0
Spot price	128.0	23209.6	6988.9	12440.0	16405.0	23020.0	28195.0	36390.0
Previous closing price	128.0	24502.2	3731.9	16735.0	20533.8	25500.0	27297.5	29380.0

TABLE 5: Parameters setting.

Model	Parameters setting
CEEMDAN	Trials = 100; epsilon = 0.005
EEMD	Trials = 100; noise_width = 0.05
WOA-LightGBM	n_estimators = 18; learning_rate = 0.027; max_depth = 10; num_leaves = 9
WOA-XGBoost	n_estimators = 38; learning_rate = 0.064; colsample_bytree = 0.3015; max_depth = 54; subsample = 0.2258
WOA-GBDT	n_estimators = 10; learning_rate = 0.1002; min_samples_leaf = 10; max_depth = 12; min_samples_split = 4
ELM	Number of nodes in the hidden layer: 10
BPNN	Number of nodes in the hidden layer: 8
SVR	C = 40; gamma = 36
WOA	Number of iterations: 100; population size: 10
GWO	Number of iterations: 100; population size: 10

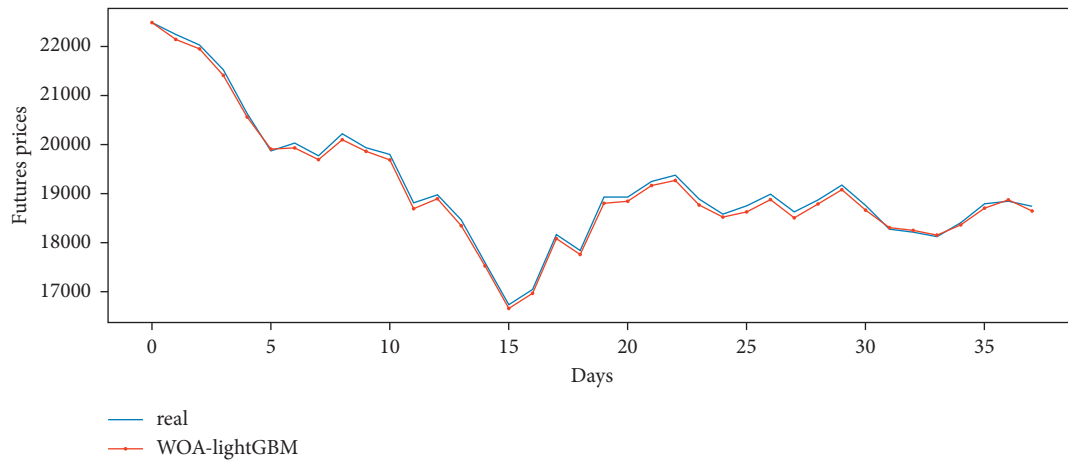


FIGURE 3: Preliminary fitting curve of WOA-LightGBM model.

has the largest prediction error in Table 7. (b) The modal decomposition algorithm allows the signal to be decomposed in accordance with the residual series' own time scale, which holds a significant advantage in dealing with non-linear and nonstationary data. Thus, the residual correction models predicted after the CEEMDAN and EEMD decomposition in the table are all superior to the residual correction models predicted directly using the SVR model. (c) Using the modal decomposition algorithm, it is further demonstrated that the CEEMDAN method by adding adaptive white noise can reduce the reconstruction error generated during the decomposition process and, as a result, improve the prediction accuracy by comparing the residual correction effects of CEEMDAN and EEMD.

4.4. Experiments in Other Financial Data. In order to test the predictive power of this paper's model in other areas, the daily trading data of the CSI 300 stock index futures are used. The data was obtained from the Wind database. The data selected spans from January 4, 2016, to July 21, 2021, with a total of 1,351 transactions. It is specifically selected as the forecast target, and the opening price, high price, low price, volume, raising limit price, and limit down price are taken as the influencing factors for the prediction. 70% of the data will be used as a training set and 30% as a testing set. Table 8 provides the calculation results of the six indicators for the model in this paper and the comparison models.

The results of the indicator analysis in Table 8 lead to the following conclusions. First of all, among the single machine

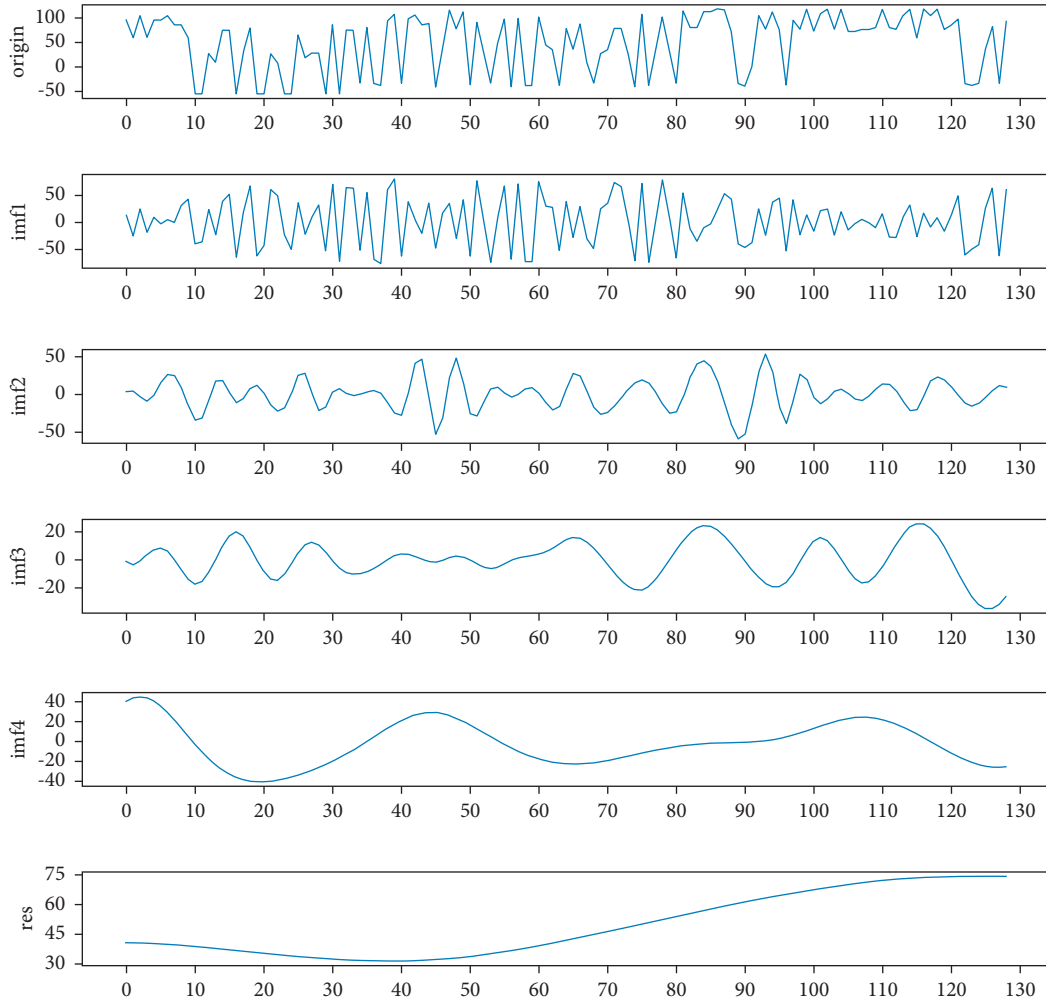


FIGURE 4: CEEMDAN decomposition of WOA-LightGBM prediction residuals.

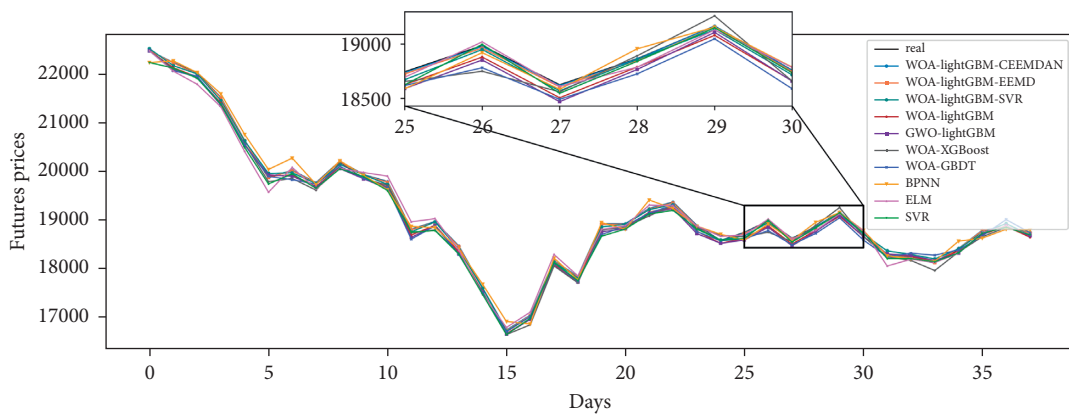


FIGURE 5: Predicted performance of the test set.

learning models, the LightGBM model optimized by the WOA algorithm has the best accuracy. In addition, its prediction metrics of MSE, MAE, RMSPE, MAPE, R^2 , and DA improved by 34.89%, 22.46%, 19.82%, 23.01%, 0.01%, and 0.52%, respectively, compared with the GWO-

LightGBM model, which ranked second among the single models. Second, all three residual correction combinations outperform the single machine learning model, suggesting that further forecasting of the residual series and the extraction of information may enhance the prediction

TABLE 6: Forecast accuracy table for hog futures.

	MSE	MAE	RMSPE (%)	MAPE (%)	R ²	DA
WOA-LightGBM-CEEMDAN	58.3171	5.2654	0.0407	0.0277	99.9964e-02	0.9737
WOA-LightGBM-EEMD	154.3967	11.3160	0.0653	0.0595	99.9905e-02	0.9737
WOA-LightGBM-SVR	2540.5649	42.5720	0.2638	0.2223	99.8432e-02	0.9474
WOA-LightGBM	7465.1947	81.1565	0.4512	0.4246	99.5393e-02	0.9737
GWO-LightGBM	11895.0375	100.7162	0.5678	0.5262	99.2659e-02	0.9211
WOA-XGBoost	11987.5541	95.6370	0.5793	0.5036	99.2602e-02	0.9211
WOA-GBDT	14009.2782	100.5755	0.6179	0.5251	99.1353e-02	0.8947
BPNN	12584.6199	88.6895	0.5844	0.4641	99.2233e-02	0.9474
ELM	13074.8747	89.1647	0.5781	0.4580	99.1931e-02	0.9737
SVR	13390.6722	89.7970	0.5905	0.4616	99.1736e-02	0.9474

Bold shows the best predicted results in the classification of each indicator.

TABLE 7: Fitting error analysis of the four models.

Date	Real	CEEMDAN		EEMD		SVR		WOA-LightGBM	
		Fitted value	Error	Fitted value	Error	Fitted value	Error	Fitted value	Error
2021/5/28	22485	22489.67	-4.67	22484.88	0.12	22530.28	-45.28	22485	0
2021/5/31	22245	22242.76	2.24	22237.85	7.15	22180.4	64.60	22140.26	104.74
2021/6/1	22025	22015.68	9.32	22008.5	16.50	22011.98	13.02	21948.65	76.35
2021/6/2	21525	21523.93	1.07	21518.05	6.95	21463.73	61.27	21412.14	112.86
2021/6/3	20640	20642.38	-2.38	20637.71	2.29	20630.58	9.42	20563.5	76.50
2021/6/4	19870	19883.66	-13.66	19878.53	-8.53	19960.14	-90.14	19908.5	-38.50
2021/6/7	20025	20024.6	0.40	20017.03	7.97	19974.63	50.37	19929.61	95.39
2021/6/8	19770	19767.81	2.19	19759.64	10.36	19752.83	17.17	19693.65	76.35
2021/6/9	20215	20211.61	3.39	20201.77	13.23	20148.2	66.80	20096.61	118.39
2021/6/10	19935	19932.91	2.09	19921.95	13.05	19932.15	2.86	19862.48	72.52
...
2021/7/8	18865	18862.46	2.54	18849.55	15.45	18858.18	6.82	18788.5	76.50
2021/7/9	19165	19164.54	0.46	19151.03	13.97	19131.55	33.45	19079.9	85.10
2021/7/12	18760	18758.54	1.46	18748.36	11.64	18717.02	42.98	18662.11	97.89
2021/7/13	18270	18284.68	-14.68	18272.09	-2.09	18364.79	-94.79	18304.53	-34.53
2021/7/14	18210	18225.68	-15.68	18216.4	-6.40	18292.82	-82.82	18248.5	-38.50
2021/7/15	18120	18137.65	-17.65	18128.5	-8.50	18199.54	-79.53	18154.53	-34.53
2021/7/16	18400	18415.71	-15.71	18408.92	-8.92	18407.64	-7.64	18363.31	36.69
2021/7/19	18785	18793.16	-8.16	18786.45	-1.45	18743.8	41.20	18702.32	82.68
2021/7/20	18840	18862.93	-22.93	18855.92	-15.92	18928.48	-88.48	18874.53	-34.53
2021/7/21	18740	18743.18	-3.18	18735.75	4.25	18691.71	48.29	18647.38	92.62
Average error		-1.35		8.59		16.28		71.65	

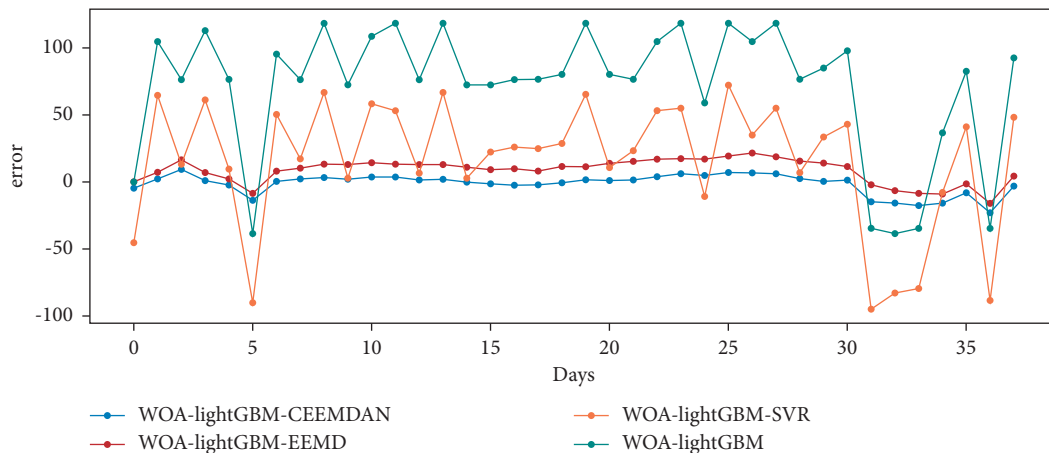


FIGURE 6: Model prediction error analysis.

TABLE 8: Forecast accuracy table of CSI 300 stock index futures.

	MSE	MAE	RMSPE (%)	MAPE (%)	R ²	DA
WOA-LightGBM-CEEMDAN	9.5274	2.3375	0.0668	0.0509	99.9971e-02	0.9753
WOA-LightGBM-EEMD	13.7575	2.7583	0.0793	0.0595	99.9958e-02	0.9704
WOA-LightGBM-SVR	28.9593	3.9951	0.1187	0.0872	99.9911e-02	0.9580
WOA-LightGBM	56.6955	4.9665	0.1610	0.1064	99.9827e-02	0.9432
GWO-LightGBM	87.0789	6.4052	0.2008	0.1382	99.9734e-02	0.9383
WOA-XGBoost	103.8484	7.9996	0.2242	0.1759	99.9682e-02	0.9333
WOA-GBDT	197.0538	10.3638	0.3015	0.2249	99.9397e-02	0.9259
ELM	166.7355	9.0973	0.2846	0.1994	99.9489e-02	0.9358
BPNN	264.4969	11.5221	0.3634	0.2512	99.9191e-02	0.9037
SVR	354.7475	10.4587	0.4366	0.2356	99.8915e-02	0.9136

Bold shows the best predicted results in the classification of each indicator.

TABLE 9: Ranking the features importance of hog futures prices.

LightGBM		XGBoost		GBDT	
Features	Score	Features	Score	Features	Score
Volume	0.354167	Sow price	0.203349	Sow price	0.205162
Sow price	0.270833	Lamb price	0.129648	Volume	0.204963
Previous closing price	0.229167	Beef price	0.12844	Corn price	0.189453
Piglet price	0.0625	Spot price	0.121461	Spot price	0.167043
Spot price	0.0625	Piglet price	0.115235	Piglet price	0.109867
Lamb price	0.020833	Corn price	0.114087	Previous closing price	0.065497
Corn price	0	Previous closing price	0.104195	Lamb price	0.035748
Beef price	0	Volume	0.083584	Beef price	0.022267

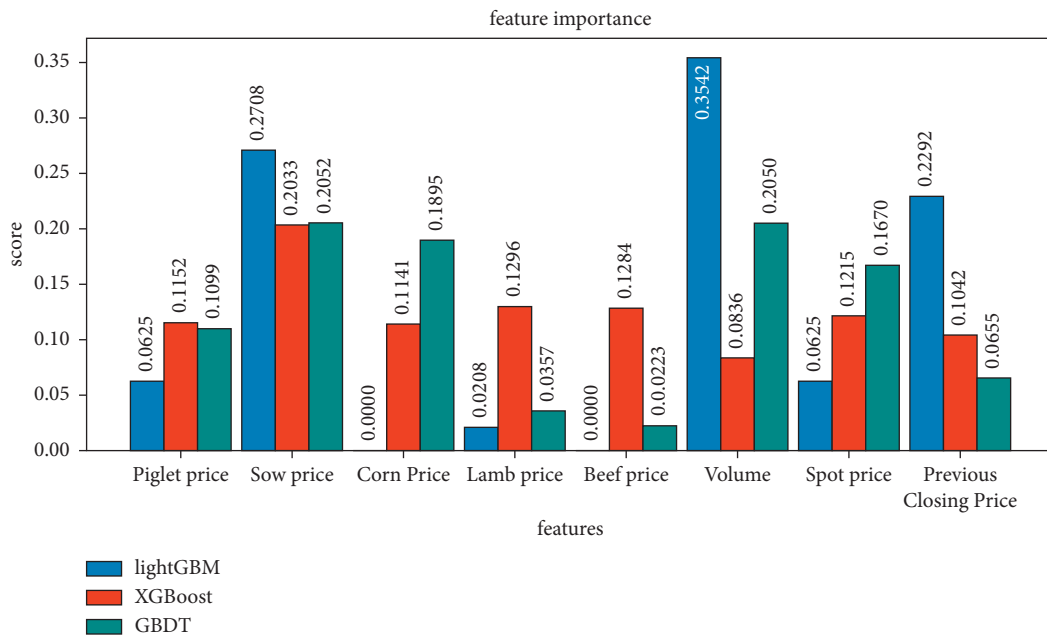


FIGURE 7: Histogram of feature importance.

accuracy of the model. Last but not least, in the three combined prediction models, WOA-LightGBM-CEEMDAN showed the best prediction results. Compared to the second-ranked WOA-LightGBM-EEMD model, the calculation results of the six prediction indexes were improved by 30.75%, 15.26%, 15.7%, 14.45%, 0.001%, and 0.50%,

respectively, indicating that the correction method of decomposition-individual forecasting-ensemble residual series by CEEMDAN, which is applied in this study, can be effective in extracting valuable information from prediction residuals of LightGBM model, improving the model's predictions.

4.5. Analysis of Feature Importance. In order to determine the most influential factors in hog futures price forecasting, the WOA-LightGBM, WOA-XGBoost, and WOA-GBDT models are used to analyze the feature importance. Table 9 and Figure 7 provide the results of the feature importance ranking for each model. The preliminary observations indicate that the results of the feature ranking in the three models are not exactly the same. The further analysis of the feature rankings in Table 9 reveals that the sow price, piglet price, and spot price are ranked in the top five out of the three models that are essential in forecasting hog futures prices. Specifically, first of all, the change in sow price reflects the market's current replenishment sentiment and influences the number of hogs that will be slaughtered after six months. A high sow price indicates positive market sentiment and a high pig slaughter volume after six months. At the same time, the price of sows is also an important part of the cost of hog farming. Secondly, the piglet price change can also reflect the market's eagerness for replenishment, reflecting the scale of pig slaughter in the past four months, and the piglet price with a lag of four months has a direct correlation with the current hog price. Finally, spot prices are inherently correlated with futures prices, particularly when they are close to delivery.

5. Conclusions

In this paper, a combined forecasting model of hog futures prices is developed by using WOA-LightGBM and CEEMDAN, which addresses the shortcomings of single machine learning models in terms of forecasting accuracy and model stability. In the first place, we define the index system of hog futures price influencing factors from three perspectives: supply, demand, and futures market, and we use grey correlation analysis to screen the indexes. Secondly, we decompose and correct the prediction residual series of the WOA-LightGBM model using the CEEMDAN method in order to construct a combined WOA-LightGBM-CEEMDAN prediction model. The following conclusions were reached as a result of simulation experiments on hog futures price data.

Firstly, the decomposition and correction of the residual sequences generated by the WOA-LightGBM model using the CEEMDAN method can enhance the prediction accuracy of the model compared to a single machine learning model. In addition, the CEEMDAN method can extract the effective information of residual series at different frequency scales in the residual correction combination model compared with the EEMD and SVR models, thus improving the model prediction accuracy. Secondly, by applying this paper's model to the prediction problem of CSI 300 stock index futures prices, it is discovered that the combined model presented in this paper offers the highest level of prediction accuracy when compared with the comparison algorithm, indicating the applicability of this paper's model to other prediction problems. Finally, we used three machine learning models, LightGBM, XGBoost, and GBDT, to model and perform feature importance analyses on hog futures prices separately and found that sow price, piglet price, and

spot price are the most influential factors when predicting hog futures prices.

In light of the above research conclusions and the methods recommended in this paper, the following suggestions are made. To begin with, the CEEMDAN method employs the concept of "decomposition-individual forecasting-ensemble" to correct the residual series, which has the potential to improve the prediction accuracy of the model and establish a new method for researchers to use in future research projects on the prediction of financial data such as futures and stocks. Secondly, the combined model proposed in this paper can more accurately predict the future price trend compared with the single machine learning model, and investors are able to receive reference support for future investment decisions based on the forecasting results of the model. Finally, the LightGBM model is able to rank the features of the model as a function of the prediction process, thereby determining the most influential factors in the model's performance. Thus, for market regulators, the feature importance analysis of the LightGBM model can be used to adjust prices of factors that have the greatest influence on futures prices, thus achieving a regulation of futures prices.

It is important to note, however, that, in this paper, the time span of the hog futures prices examined is short, and the data sample is small, so the time series information of the hog futures prices is not fully utilized in building the model. Also, due to the short time span of the selected data, the cycle effect of hog futures prices was not taken into account when the data was processed. As such, future research should consider the cycle effects of hog futures prices, and the time series information should also be integrated into the forecasting process of the model in order to improve its forecasting performance.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare no conflicts of interest.

Acknowledgments

This research was funded by the National Natural Science Foundation of China (Grant no. 81973791).

References

- [1] C. G. Sheng, G. Y. Wang, Y. D. Geng, and L. R. Chen, "The correlation analysis of futures pricing mechanism in China's carbon financial market," *Sustainability*, vol. 20, no. 18, 2020.
- [2] C. Y. Wang, "Forecast on price of agricultural futures in China based on arima model," *Asian Agricultural Research*, vol. 8, no. 11, pp. 9–12, 2016.
- [3] A. Trujillo-barrera, P. Garcia, and M. L. Mallory, "Short-term price density forecasts in the lean hog futures market,"

- European Review of Agricultural Economics*, vol. 45, no. 1, pp. 121–142, 2018.
- [4] O. V. De la torre-torres, D. Aguila-socho-montoya, and M. D. del rio-rama, “A two-regime markov-switching garch active trading algorithm for coffee, cocoa, and sugar futures,” *Mathematics*, vol. 8, no. 6, 2020.
 - [5] Q. H. Gu, Y. X. Chang, N. X. Xiong, and L. Chen, “Forecasting nickel futures price based on the empirical wavelet transform and gradient boosting decision trees,” *Applied Soft Computing*, vol. 109, 2021.
 - [6] J. Chai, Y. Wang, S. Wang, and Y. Wang, “A decomposition-integration model with dynamic fuzzy reconstruction for crude oil price prediction and the implications for sustainable development,” *Journal of Cleaner Production*, vol. 229, pp. 775–786, 2019.
 - [7] X. Wang, S. Gao, S. Y. Zhou, Y. B. Guo, Y. H. Duan, and D. Q. Wu, “Prediction of house price index based on bagging integrated WOA-SVR model,” *Mathematical Problems in Engineering*, vol. 2021, Article ID 3744320, 15 pages, 2021.
 - [8] W. P. Liu, C. Z. Wang, Y. G. Li, Y. S. Liu, and K. K. Huang, “Ensemble forecasting for product futures prices using variational mode decomposition and artificial neural networks,” *Chaos, Solitons & Fractals*, vol. 146, 2021.
 - [9] J. Manickavasagam, S. Visalakshmi, and N. Apergis, “A novel hybrid approach to forecast crude oil futures using intraday data,” *Technological Forecasting and Social Change*, vol. 158, Article ID 120126, 2020.
 - [10] J. H. Chen, S. F. Ma, and Y. Wu, “International carbon financial market prediction using particle swarm optimization and support vector machine,” *Journal of Ambient Intelligence and Humanized Computing*, vol. 2021, pp. 1–15, 2021.
 - [11] R. Y. Chen, B. Pan, and X. D. Lin, “Chinese stock index futures price fluctuation analysis and prediction based on complementary ensemble empirical mode decomposition,” *Mathematical Problems in Engineering*, vol. 2016, Article ID 3791504, 13 pages, 2016.
 - [12] J. Wang, J. Wang, and N. Zhang, “A new hybrid forecasting model based on SW-LSTM and wavelet packet decomposition: a case study of oil futures prices,” *Computational Intelligence and Neuroscience*, vol. 2021, Article ID 7653091, 22 pages, 2021.
 - [13] Y. L. Zhang and S. Hamori, “Forecasting crude oil market crashes using machine learning technologies,” *Energies*, vol. 13, no. 10, 2020.
 - [14] S. K. Deng, X. R. Huang, Z. H. Qin, F. Zhe, and T. X. Yang, “A novel hybrid method for direction forecasting and trading of apple futures,” *Applied Soft Computing*, vol. 110, 2021.
 - [15] G. H. Luo, J. Chen, and X. M. Ke, “Research on the simulation of futures commodity transaction price forecast,” *Computer Simulation*, vol. 35, no. 9, pp. 421–426, 2018.
 - [16] Y. Zhang, J. Chen, X. F. Wang, and F. Liu, “Application of xgboost to fault diagnosis of rolling bearings,” *Noise and Vibration Control*, vol. 37, no. 4, pp. 166–170, 2017.
 - [17] D. Li, J. Peng, and D. He, “Aero-engine exhaust gas temperature prediction based on lightgbm optimized by improved bat algorithm,” *Thermal Science*, vol. 22, no. 1, 2021.
 - [18] J. Yan, Y. T. Xu, Q. Cheng et al., “Lightgbm: accelerated genomically designed crop breeding through ensemble learning,” *Genome Biology*, vol. 22, no. 1, 2021.
 - [19] Y. Ju, G. Sun, Q. Chen, M. Zhang, H. Zhu, and M. U. Rehman, “A model combining convolutional neural network and lightgbm algorithm for ultra-short-term wind power forecasting,” *IEEE Access*, vol. 7, pp. 28309–28318, 2019.
 - [20] L. Zhang, Y. Dong, and J. Wang, “Wind speed forecasting using a two-stage forecasting system with an error correcting and nonlinear ensemble strategy,” *IEEE Access*, vol. 7, pp. 176000–176023, 2019.
 - [21] Y. Li, Z. J. Qian, W. Deng, D. Z. Zhang, H. H. Lu, and S. H. Wang, “Forecasting crude oil prices based on variational mode decomposition and random sparse bayesian learning,” *Applied Soft Computing*, vol. 113, 2021.
 - [22] M. Niu, Y. Hu, S. Sun, and Y. Liu, “A novel hybrid decomposition-ensemble model based on VMD and HGWO for container throughput forecasting,” *Applied Mathematical Modelling*, vol. 57, pp. 163–178, 2018.
 - [23] H. Liu and L. Shen, “Forecasting carbon price using empirical wavelet transform and gated recurrent unit neural network,” *Carbon Management*, vol. 11, no. 1, pp. 25–37, 2020.
 - [24] N. E. Huang, Z. Shen, S. R. Long et al., “The Empirical Mode Decomposition and the Hilbert Spectrum for Nonlinear and Non-stationary Time Series Analysis,” *Proceedings: Mathematical, Physical and Engineering Sciences*, vol. 454, no. 1971, pp. 903–995, 1998.
 - [25] Z. H. Wu and N. E. Huang, “Ensemble empirical mode decomposition: A noise — assisted data analysis method,” *Advances in Adaptive Data Analysis*, vol. 1, no. 1, pp. 1–41, 2009.
 - [26] M. E. Torres and M. A. Colominas, “A complete ensemble empirical mode decomposition with adaptive noise,” in *Proceedings of the 2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 4144–4147, Prague, Czech Republic, May 2011.
 - [27] W. Zhang, Z. Qu, K. Zhang, W. Mao, Y. Ma, and X. Fan, “A combined model based on CEEMDAN and modified flower pollination algorithm for wind speed forecasting,” *Energy Conversion and Management*, vol. 136, pp. 439–451, 2017.
 - [28] J. Wang, J. X. Cao, S. Yuan, and M. Cheng, “Short-term forecasting of natural gas prices by using a novel hybrid method based on a combination of the CEEMDAN-SE-and the PSO-ALS-optimized GRU network,” *Energy*, vol. 233, 2021.
 - [29] J. Zhou and D. Chen, “Carbon price forecasting based on improved CEEMDAN and extreme learning machine optimized by sparrow search algorithm,” *Sustainability*, vol. 13, no. 9, 2021.
 - [30] J. Cao, Z. Li, and J. Li, “Financial time series forecasting model based on CEEMDAN and LSTM,” *Physica A: Statistical Mechanics and Its Applications*, vol. 519, pp. 127–139, 2019.
 - [31] CBS, *China Statistical Yearbook*, China Statistics Press, Beijing, China, 2020.
 - [32] A. Sharma and B. Singh, “AE-LGBM: Sequence-based novel approach to detect interacting protein pairs via ensemble of autoencoder and LightGBM,” *Computers in Biology and Medicine*, vol. 125, 2020.
 - [33] S. Mirjalili and A. Lewis, “The whale optimization algorithm,” *Advances in Engineering Software*, vol. 95, pp. 51–67, 2016.
 - [34] W. Z. Liang, S. Z. Luo, G. Y. Zhao, and H. Wu, “Predicting hard rock pillar stability using GBDT, XGBoost, and LightGBM algorithms,” *Mathematics*, vol. 8, no. 5, 2020.
 - [35] H. B. Luo and D. M. You, “Integrated prediction on stock index series based on main trend identification and intelligent error correction,” *Systems Engineering*, vol. 33, no. 2, pp. 24–30, 2015.
 - [36] Y. S. Zhu, W. H. Wang, and B. T. Han, “Grey correlative degree analysis of factors affecting demands in the real estate market,” *Transactions of Beijing Institute of Technology*, vol. 6, pp. 782–785, 2002.
 - [37] Y. Zhao, J. Li, and L. Yu, “A deep learning ensemble approach for crude oil price forecasting,” *Energy Economics*, vol. 66, pp. 9–16, 2016.